Renewable Energy and Power Quality Journal
https://repqj.com/index.php/repqj/index
RE&PQJ, Vol. 23, No. 6, 2025

Reinforcement Learning Algorithm Optimizes Multi-Sensor Energy Management Strategy for New Energy Vehicles

Shanshan Li*

Henan Quality Institute, Pingdingshan, 467000, Henan Province, China
*Corresponding author's email: lishanshan016@126.com

Abstract. Aiming at the problem that the current multi-sensor Energy Management Strategy (EMS) of new energy vehicles is not adaptable enough and has insufficient system stability when dealing with sensor data mutations and complex road conditions, this paper takes Plug-in Hybrid Electric Vehicle (PHEV) as the research object, integrates Reinforcement Learning (RL) algorithm, and studies the optimization of multi-sensor EMS, aiming to improve the energy consumption control and system robustness of PHEV under non-steady-state conditions and sensor interference conditions. This paper first constructs a state observation module that integrates multi-sensor data to provide input for decision-making strategies through refined perception of the vehicle's operating environment. Then, a dual network structure is introduced based on Deep Q-Network (DQN) to alleviate the problem of Q-value overestimation and improve strategy stability by separating action selection and value evaluation processes. Finally, combined with Prioritized Experience Replay (PER), the training priority of experience samples is dynamically adjusted according to the Temporal Difference (TD) error to improve the learning efficiency of key states and the generalization ability of strategies. The conclusion shows that the EMS under the proposed method has strong dynamic load stability and adaptability in complex working conditions of a disturbance environment, providing a new idea with more engineering adaptability for the energy efficiency optimization of PHEV.

Key words. Energy management strategy, Multi-sensor fusion, Plug-in hybrid electric vehicle, Double Deep Q-Network, Priority experience replay

List of abbreviations/Symbols

Sequence	Abbreviations /Symbols	Full name/definition
1	EMS	Energy Management Strategies
2	PHEV	Plug-in Hybrid Electric Vehicle
3	RL	Reinforcement Learning
4	DQN	Deep Q-Network
5	PER	Prioritized Experience Replay
6	TD	Temporal Difference
7	SOC	State of Charge
8	EV	Electric vehicle
9	DRL	Deep Reinforcement Learning
10	LSTM	Long Short-Term Memory
11	RC	Resistance-Capacitance
12	SAC	Soft Actor-Critic
13	PPO	Policy Optimization
14	T_{batt}	Battery temperature
15	n_{motor}	Motor speed
16	n_{eng}	Engine speed
17	v_{veh}	Vehicle speed
18	θ_{acc}	Accelerator pedal opening
19	θ_{brake}	Brake pedal opening
20	ϕ_{road}	Road slope
21	T_{env}	Ambient temperature
22	Δt	Global unified sampling period
23	t_k	Resampling node
24	N	Window size
25	x_k	States
26	F	State transition matrix
27	H	Observation matrix
28	ω_k	Process noise
29	v_k	Observing noise
30	m	Quality
31	v	Speed
32	a	Acceleration
33	F_{dive}	Drive
34	μ	Coefficient of rolling resistance
35	Air_ρ	Air density
36	Air_A	Air windward area
37	Air_d	Air resistance coefficient
38	ϕ	Road inclination angle
39	U_{oc}	Open circuit voltage
40	R_o	Ohmic resistance
41	Q_{bat}	Rated capacity of battery

42	η_{bat}	Charge/discharge efficiency
43	T_{eng}	Torque
44	n_{eng}	Rotational speed
45	ρ_{fuel}	Energy density per volume of fuel
46	i_{eng}	Instantaneous thermal efficiency
47	η_{gen}	Motor Efficiency
48	R_t	Total reward function
49	S_t	State Space
50	A_t	ACTION SPACE
51	γ	Discount factor
52	Q_{online}	Discount factor
53	Q_{target}	Target Value Network
54	δ_i	TD error
55	ς	Factors determining priority order
56	ω_i	Importance sampling weight

1. Introduction

In the current world where energy crisis and environmental problems are becoming increasingly serious, PHEV, as a new type of transportation that can replace traditional fuel vehicles, has become a core technology to promote the development of transportation towards green direction [1,2]. PHEV multi-sensor EMS is an important factor affecting vehicle energy efficiency, endurance, and user experience, and determines the vehicle performance level [3,4]. To achieve effective coordination between power batteries, energy recovery, electric drive, and other parts, multiple types of sensors are installed in the vehicle to provide support for energy management by real-time monitoring of important parameters such as battery power, vehicle speed, road slope, ambient temperature, etc. [5,6]. The current EMS has improved the energy-saving performance of PHEV to a certain extent, but because it is too dependent on pre-set models, it is difficult to effectively adjust to sudden sensor data and driving environment, resulting in poor robustness and real-time performance in practical applications. These problems seriously limit the energy efficiency optimization potential of PHEV in complex road environments.

RL algorithm is an important branch of machine learning. It learns through the interaction between the subject and the environment, thereby achieving individual adaptability and dynamic optimization [7,8]. This paper enhances the perception ability of vehicle operating environment by integrating a multi-sensor data fusion module, enabling strategies to more accurately respond to complex and nonlinear driving needs; compared with previous studies, the "Dual DQN+PER" framework in this paper has been optimized specifically for sensor data mutations and complex road conditions. By applying a dual network structure to separate the process of action selection and value evaluation, the problem of overestimation of Q-values can be reduced; based on the PER mechanism, the importance of samples is dynamically adjusted according to TD error, making the model more focused on learning key state transitions, which is particularly important for dealing with sensor noise and rapidly changing road conditions.

With the development of new energy vehicle technology, existing research is devoted to solving problems such as power distribution and battery utilization in vehicle energy management [9,10]. To achieve electric vehicle (EV) EMS optimization, Kranthikumar et al. proposed a combination of a bidirectional long short-term memory network based on enhanced multi-head cross attention and the Remora optimization algorithm. They implemented the proposed method in Matlab and compared it with several other benchmarks. The results showed that the regenerative braking efficiency using the proposed technology was reduced to 4.5% [11]. Hong et al. proposed an EV real-time EMS strategy based on deep Long Short-Term Memory (LSTM) to manage large-scale EVs in layers and partitions. They used historical load information to obtain the historical optimal solution to train the learning network to guide new real-time optimization. Finally, he verified the effectiveness and superiority of the proposed layered architecture and management strategy through simulation examples [12]. To solve the problem of battery capacity attenuation caused by excessive battery discharge current during EV driving, An et al. used fuzzy logic controller to adjust the charge and discharge power of lithium-ion power batteries and supercapacitors based on fuzzy logic EMS and 45 established fuzzy control rules. The experimental results under test conditions showed that the proposed strategy could effectively avoid the influence of current fluctuations and extend the battery life [13]. Guo et al. proposed an improved low-pass filter equivalent power minimization EMS, which smoothed the transient changes of fuel cell power by applying low-pass filtering technology. Simulation results showed that the proposed strategy performed well in suppressing battery power fluctuations under idling conditions and significantly improved the operating efficiency of the battery [14]. These studies provide certain guidance for improving automobile energy efficiency and power distribution, but there are still limitations in long-term performance optimization and overall economy under dynamic environments.

RL can dynamically perceive multi-sensor information through interactive learning between the agent and the environment, providing more possibilities for improving the long-term performance and overall economy of EMS [15,16]. Li et al. proposed a method based on a solid fuel cell model and an advanced DRL (Deep Reinforcement Learning) algorithm, supplemented by expert knowledge of rule-based EMS. The proposed method was thoroughly tested in various scenarios. The results showed that the proposed method was superior to existing methods in terms of long-term learning efficiency and improved driving economy by 2.8% to 7.5% [17]. Zou et al. used a range-extended vehicle as the research object to explore the optimal EMS based on rules and the optimal EMS based on RL for the vehicle, and built a strategy model for simulation in MATLAB software. The results showed that the energy consumption rate of the EMS optimized based on RL was 3.2% lower than that of the original rule-based EMS [18]. To solve the problem of low energy saving effect of hybrid electric vehicle EMS when running online, Chen et al. proposed a DRL-based EMS design method. The established control strategy included a two-layer logic framework of offline interactive learning and online update learning, and dynamically updated the control parameters according to the vehicle operation characteristics. The results showed that the proposed DRL-based EMS could achieve long-term energy saving effect better than the particle swarm optimization strategy [19]. Tang et al. proposed an EMS based on a deep value network algorithm. Through multi-objective collaboration, it realized the upper-level vehicle following control and lower-level energy management for PHEVs. Finally, through simulation, it was verified that the DRL-based EMS achieved good fuel economy in both the pilot vehicle and the following vehicle [20]. These studies have promoted the development of EMS for new energy vehicles in the direction of green and economic efficiency. However, most current RL applications need to further improve their strategy stability and generalization capabilities when faced with sudden sensor data changes and complex road conditions.

To improve the energy consumption control capability and system robustness of PHEV in unstable environments, this paper combines RL algorithms to study the optimization of PHEV multi-sensor EMS. Innovations: 1) this paper integrates a multi-sensor data fusion module to enhance the perception of the vehicle's operating environment, enabling the strategy to respond more accurately to complex and nonlinear driving needs; 2) this paper applies a dual DQN structure, which effectively alleviates the policy deviation problem caused by the overestimation of Q values in traditional DQN by separating the action selection and action evaluation processes, thereby improving the algorithm’s learning stability and policy accuracy in complex state spaces; 3) the PER mechanism is integrated to dynamically adjust the sampling probability of the empirical samples according to the TD error, so that the model pays more attention to the key state transition, significantly improving the sample utilization efficiency and convergence speed, and providing a more adaptive thinking for EMS deployed in real scenarios.

2. Construction of Multi-sensor EMS for New Energy Vehicles

A. System Modeling

This paper takes PHEV as the research object. PHEV has both traditional engine and electric motor energy, including battery energy management, engine start-stop control, energy recovery scheduling, and multiple decision-making issues. The energy flow is complex and highly dependent on sensor data, which can fully reflect the fusion management characteristics of multi-sensor information [21-23]. In addition, the energy management of PHEV also faces complex and changeable driving conditions and environmental changes, and needs to have strong adaptability, real-time, and robustness. Based on the PHEV architecture, this paper can model its multi-sensor energy management system.

1) Sensor Data Integration

In system modeling, the important sensor information related to energy management is first integrated. To ensure the real-time, stable, and accurate performance of the vehicle under complex driving conditions, this paper uses the four types of sensor data that are most widely involved in energy management as the basis for state space input and analyzes them, as shown in Table 1.

Table 1. Sensor data.

Parameter	Classification	Variables
Battery status	SOC	SOC
	Battery temperature	T_{batt}
Power system operating	Motor speed	n_{motor}
	Engine speed	n_{eng}
Driving behavior	Vehicle speed	v_{veh}
	Accelerator pedal opening	θ_{acc}
	Brake pedal opening	θ_{brake}
External environmental	Road slope	ϕ_{road}
	Ambient temperature	T_{env}

Firstly, to solve the problem that the acquisition timing of multiple sensors is inconsistent, and there is a deviation in the timing, a synchronization method based on a unified time axis is used to process it. The global unified sampling period Δt is set, and the nearest neighbor interpolation and linear interpolation are used to reconstruct the asynchronous data. The original sensor data sequence is set to $\left\{x_i\left(t_i\right)\right\}$, and the estimation at the uniform resampling node $t_k = k\Delta t$ is calculated by the formula:

$$x\left(t_k\right)=x\left(t_i\right)+\frac{x\left(t_{i+1}\right)-x\left(t_i\right)}{t_{i+1}-t_i}\times\left(t_k-t_i\right)\quad\left(t_i\le t_k\le t_{i+1}\right)\quad(1)$$

For data such as θ_{acc} and θ_{brake}, since their signals fluctuate greatly and are prone to short-term mutations, the sliding average method is used synchronously during interpolation, and the window size N is dynamically adjusted according to the rate of change of the signal, taking 5 to 10 frames. The calculation formula is [24]:

$$\hat{x}\left(t_k\right)=\frac{1}{N}\sum_{j=0}^{N-1}x\left(t_{k-j}\right)\quad(2)$$

Synchronization and smoothing are used to make each sensor signal output at the same time, constructing input features with continuity and consistency. The dynamic estimation algorithm based on the extended Kalman Filter reduces the interference of sensor noise and improves the accuracy of the system's state estimation. The true value of each sensor is taken as the state vector x_k of the system; the sensor measurement value is taken as the observation vector z_k; the state transition and observation equations are modeled:

$$x_{k+1}=Fx_k+\omega_k\quad(3)$$

$$z_k=Hx_k+v_k\quad(4)$$

Here, F represents the state transition matrix; H represents the observation matrix; ω_k and v_k represent the process noise and observation noise, and they are assumed to be Gaussian white noise with a mean of 0. To solve the problem of inconsistent acquisition time of different sensors, the unified time axis synchronization method is first applied, and then, the asynchronous data is reconstructed using the nearest neighbor interpolation. Next, combined with extended Kalman filter, the state prediction for the next time is adjusted based on the current state estimation and measurement values in each update step, effectively reducing errors caused by asynchronous sensor sampling.

In each update step, the extended Kalman filter adjusts the next state prediction $\hat{x}_{k-1|k-1}$ based on the current state estimation and measurement values. Firstly, based on the previous state estimation $\hat{x}_{k-1|k-1}$, the state transition equation is used to predict the current state:

$$\hat{x}_{k|k-1}=f\left(\hat{x}_{k-1|k-1},u_k\right)\quad(5)$$

The error covariance matrix $\mathbf{C}$ of the predicted state $P_{k|k-1}$ is calculated:

$$P_{k|k-1}=F_kP_{k-1|k-1}\quad(6)$$

Based on the predicted error covariance and observation model, the Kalman gain is calculated:

$$K_k=P_{k|k-1}H_K^T\left(H_kP_{k|k-1}H_K^T+R_K\right)^{-1}\quad(7)$$

When encountering sudden interference, extended Kalman filter quickly adjusts state estimation based on the latest observations to reduce the impact of interference on system performance. Although introducing Kalman filter increases the system's computational burden, it provides higher accuracy and robustness, especially in the presence of noise and abrupt changes. To balance this point, this paper optimizes the implementation of Kalman filter, reduces unnecessary calculation steps, and adopts an efficient matrix operation library to improve execution speed.

After completing data quality control, considering the heterogeneity of each sensor data, it is standardized. The continuous sensor data is mapped to intervals:

$$x'=\frac{x-x_{\text{min}}}{x_{\text{max}}-x_{\text{min}}}\quad(8)$$

Among them, x_{min} and x_{max} are the minimum and maximum values of the corresponding features in the sample set.

Based on the processing of a single feature, the feature splicing method is used to combine each standardized sensor signal according to a predetermined order to obtain a unified feature vector. Its structure is defined as:

$$f_t=\left[\text{SOC},T_{\text{batt}},n_{\text{motor,}}n_{\text{eng}},v_{\text{veh}},\theta_{\text{acc}},\theta_{\text{brake}},\phi_{\text{road}},T_{\text{env}}\right]^{\text{T}}\quad(9)$$

Here, T represents vector transposition.

2) *Dynamic Model*

Based on the integration of multi-source sensor data, this paper obtains the real-time observation value of the PHEV vehicle operation status. On this basis, a dynamic model of the vehicle power system is built, including the dynamic model of the vehicle movement, battery electrochemical behavior, and engine-generator coupling operation. The method of combining physical modeling and system identification is used to analyze the evolution law of the dynamic energy of the vehicle under complex working conditions, as shown in Figure 1.

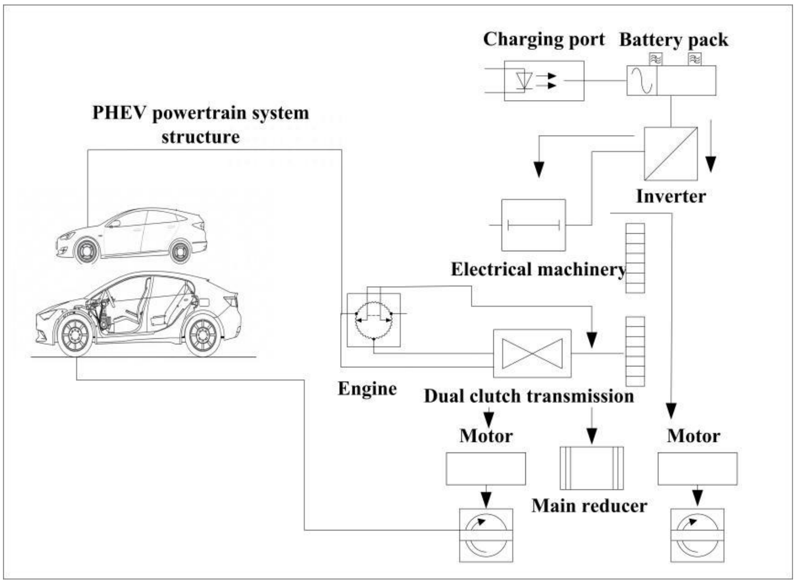

Figure 1. PHEV power system structure.

First, based on the one-dimensional driving dynamics framework of the mass point method, the vehicle dynamics model is established, taking into account the four main forces: driving force, rolling resistance, air resistance, and slope resistance. Assuming that the mass of the vehicle is m, and the speed is v, its acceleration $a = \frac{\mathrm{d}v}{\mathrm{d}t}$ is expressed by the dynamic balance equation:

$$F_{\mathrm{dive}} - F_{\mathrm{resist}} = m\frac{\mathrm{d}v}{\mathrm{d}t} \quad (10)$$

In this formula, the driving force F_{dive} determined by the output torque of the motor and engine is applied to the wheels through the transmission system; the entire F_{resist} consists of three components [25,26]:

$$F_{\mathrm{resist}} = F_{\mathrm{roll}} + F_{\mathrm{air}} + F_{\mathrm{slope}} = \mu mg\ \cos\phi + \frac{1}{2} Air_{\rho} Air_{A} Air_{d} v^{2} + mg\ \sin\phi \quad (11)$$

Among them, μ represents the rolling resistance coefficient; Air_{ρ} represents the air density; Air_{A} represents the air frontal area; Air_{d} represents the air resistance coefficient; ϕ represents the road inclination angle, which is obtained by fusing the data of multiple sensors. The model physically constrains the EMS by responding to external disturbances such as slope and wind resistance in real-time.

An improved secondary RC (Resistance-Capacitance) equivalent circuit model is used to simulate lithium-ion power batteries, taking into account the accuracy and calculation speed of the model. Considering the dynamic response of the open circuit voltage U_{oc}, the ohmic internal resistance R_{o}, and the parallel RC network (R_1, C_1 and R_2, C_2), the relationship between the battery terminal voltage U_t and the discharge current I_t is expressed as:

$$U_t = U_{\mathrm{oc}} - R_o I_t - V_{\mathrm{RC1}}(t) - V_{\mathrm{RC2}}(t) \quad (12)$$

Here, the voltages of the two RC branches satisfy the condition:

$$\frac{\mathrm{d}V_{RC_i}}{\mathrm{d}t} = -\frac{1}{R_i C_i} V_{RC_i} + \frac{1}{C_i} I_t, i = 1,2 \quad (13)$$

Then, the battery SOC state is updated in real-time using the coulomb counter measurement:

$$\mathrm{SOC}(t) = \mathrm{SOC}(t_0) - \frac{1}{Q_{\mathrm{bat}}} \int_{t_0}^{t} \eta_{\mathrm{bat}} I(\tau)\mathrm{d}\,\tau \quad (14)$$

Here, Q_{bat} represents the rated capacity of the battery, and η_{bat} represents the charge/discharge efficiency. Based on the RC dynamic response element, the terminal voltage under different charge/discharge efficiencies is estimated to solve the transient error problem caused by energy collection and accelerated conversion.

The engine-generator coupling model provides two ways of power and electricity for the hybrid system. The system's steady-state characteristic model is established by table lookup method, and the dynamic hysteresis link is used to describe the system's hysteresis characteristics. The relationship between the output torque T_{eng} and speed n_{eng} of the engine is determined by two-dimensional performance surface mapping method, and the corresponding fuel consumption rate model is given:

$$F_{\mathrm{fuel}} = f\left(n_{\mathrm{eng}}, T_{\mathrm{eng}}\right) = \frac{P_{\mathrm{eng}}}{i_{\mathrm{eng}}\left(n_{\mathrm{eng}}, T_{\mathrm{eng}}\right) \cdot \rho_{fuel}} \quad (15)$$

Among them, there is:

$$P_{\mathrm{eng}} = 2\pi n_{\mathrm{eng}} / 60 \quad (16)$$

ρ_{fuel} represents the energy density per volume of fuel, and i_{eng} represents the instantaneous thermal efficiency obtained by looking up the table. When the generator is coupled with the engine, it becomes an auxiliary power, and its output is affected by the motor efficiency η_{gen}. The maximum power limit is considered:

$$P_{\mathrm{gen}} = \eta_{\mathrm{gen}} P_{\mathrm{eng}}^{\mathrm{avail}}, \quad P_{\mathrm{gen}} \le P_{\mathrm{gen}}^{\max} \quad (17)$$

On this basis, the integrated real-time sensor state vector is used as the input driving variable, and based on the established vehicle dynamics model, battery dynamics model, and engine-generator model, a vehicle energy evolution dynamics equation group defined by state variables, system parameters, and constraints is formed.

3) *Energy Flow Model*

To characterize the energy transfer relationship between various components such as batteries, motors, engines, generators, and wheels, an energy flow map is constructed to analyze the energy flow characteristics under sudden changes in sensor data and complex working conditions, as shown in Figure 2.

In Figure 2, the energy flow diagram uses the power consumption of components under a time step as the basic description unit, as shown in Table 2:

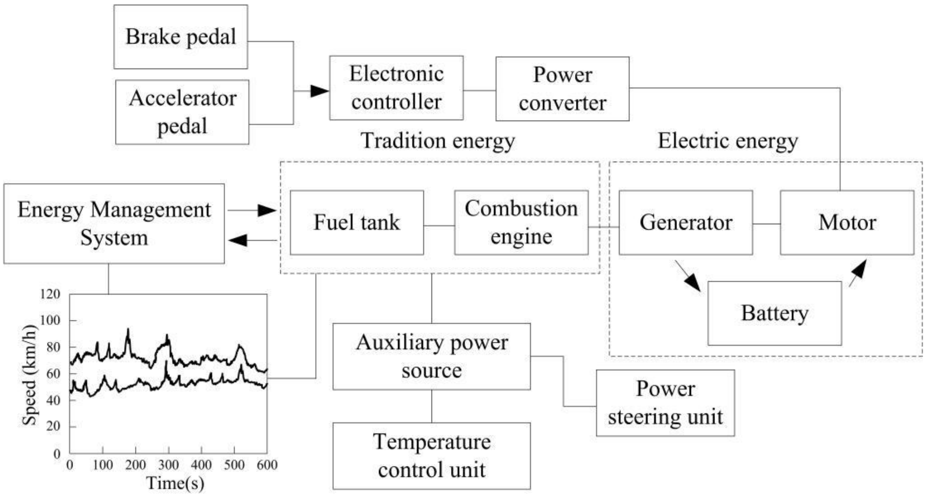

Figure 2. Energy flow diagram.

Table 2. Power status of each component.

Component	Power status	Variables
Battery	Output	$P_{\mathrm{bat}}(t)$
Traction motor	Input	$P_{\mathrm{mot}}(t)$
Engine	Output	$P_{\mathrm{eng}}(t)$
Generator	Output	$P_{\mathrm{gen}}(t)$
Wheel	Drive	$P_{\mathrm{wheel}}(t)$
Each subsystem	Energy transfer loss	$P_{loss}(t)$

According to the constructed model, the energy flow relationship of each component is defined as:

$$P_{\mathrm{wheel}}(t) = \eta_{\mathrm{mot}}(t) P_{\mathrm{mot}}(t) \quad (18)$$

$$P_{\mathrm{mot}}(t) = P_{\mathrm{bat}}(t) + \eta_{\mathrm{gen}}(t) P_{\mathrm{eng}}(t) \quad (19)$$

$$P_{\mathrm{eng}}(t) = \eta_{\mathrm{eng}}(t) P_{\mathrm{eng}}(t) \quad (20)$$

Among them, $\eta_{\mathrm{mot}}(t)$, $\eta_{\mathrm{gen}}(t)$, and $\eta_{\mathrm{eng}}(t)$ are established by table lookup and multi-point interpolation. They can be dynamically corrected according to the real-time data of the sensor, representing the instantaneous energy conversion efficiency. $P_{\mathrm{loss}}(t)$ is dispersed in each level of transmission process. According to the principle of energy conservation, there is:

$$P_{\mathrm{input}}(t) = P_{\mathrm{output}}(t) + P_{\mathrm{loss}}(t) \quad (21)$$

To more accurately describe the evolution law of energy flow state under complex operating conditions, this paper represents the energy flow in the form of node-edge based on multi-sensor fusion data:

$$G = (V, E) \quad (22)$$

Here, V represents each energy unit, and E represents the energy transfer path. The weight is used to represent the energy flow rate, that is, the instantaneous power amplitude.

The basic content of the generation process is:

(1) Initialization of V : a set of fixed nodes is determined by the dynamic model.

(2) Real-time allocation of E weights: using the power equation $P = U \times I$ and combining it with the real-time data of the sensor to calculate the energy flow between each node.

(3) Dynamic graph update: according to the vehicle's operating states such as acceleration, braking, and sliding, the edge direction and weight of (Δt) in each sample period are adjusted to achieve bidirectional energy control of the vehicle.

Considering the vehicle's complex operating conditions, the energy flow anomaly model is established with the sudden change of sensor data and road interference as external interference sources.

The abnormal energy flow caused by the sensor mutation is estimated by using the standard deviation of the power change rate in the local sliding window, and the $\sigma \Delta P > \theta_p$ condition is met. When $\sigma \Delta P > \theta_p$, the energy flow state of the system is determined to be abnormal, and the fault tolerance mechanism of the EMS is activated. Here, θ_p is the preset threshold. During the driving process of the vehicle, the vehicle's rolling resistance and slope force are corrected in real-time, and the driving power of the vehicle is transmitted to the wheel driving power $P_{\text{wheel}}(t)$, thereby directly guiding the energy distribution during the vehicle's driving process.

B. Energy Management Strategy Optimized by Reinforcement Learning

1) Deep Q-network

RL can learn optimal strategies online through interaction with the environment and adapt to complex working conditions; meanwhile, its modeling approach based on reward mechanism can directly optimize the system's overall energy efficiency target. However, traditional machine learning methods rely heavily on offline training and are difficult to cope with uncertainty in dynamic environments. Therefore, choosing reinforcement learning algorithms can better meet the real-time and adaptive requirements of new energy vehicle energy management systems. Based on system modeling, to solve the problems of sudden changes in sensor information data and dynamic instability of the environment in the system, DQN is used to build an RL framework, fit the state-action value function, and optimize the energy supply strategy. The intelligent decision-making model based on dynamic behavior interacts with the outside world, obtains reward information in real-time, and stores and uses historical experience, realizing continuous updating and smooth convergence of strategies, and enabling the automotive system to have autonomous learning and adjustment capabilities under complex operating conditions. In the high-dimensional state space, DQN constructs $Q(s,a)$ to perform dynamic learning and decision-making of EMS. It uses the data collected by sensors in real-time as the basic input and integrates the SOC, vehicle speed, acceleration, driving power, and efficiency indicators of each component to construct the current driving state vector of the vehicle and use it as the model's state space. At the same time, the action space is defined as the power distribution command between the battery and the engine to ensure that the powertrain optimizes the energy supply path while meeting driving requirements.

First, based on the integration of multi-sensor information fusion and the establishment of dynamics and energy flow models, the state space S_t of the system is defined:

$$S_t = \left\{ \text{SOC}(t), v(t), a(t), P_{\text{wheel}}(t), P_{\text{eng}}(t), P_{\text{bat}}(t), \eta_{\text{mot}}(t), \eta_{\text{gen}}(t), \eta_{\text{eng}}(t) \right\} \tag{23}$$

In this state space, the state of charge of the vehicle battery, the vehicle power demand, the output level of the engine and motor, and the efficiency parameters of the system are fully reflected and collected and updated in real-time through sensors. Among them, the $\text{SOC}(t), v(t), a(t)$ indexes come from the integrated sensor data, and the $\eta_{\text{mot}}(t), \eta_{\text{gen}}(t), \eta_{\text{eng}}(t)$ parameters are calculated in real-time through the dynamic model.

$$A_t = \left\{ P_{\text{eng}}^{\text{set}}(t), P_{\text{bat}}^{\text{set}}(t) \right\} \tag{24}$$

Among them, $P_{\text{eng}}^{\text{set}}(t)$ and $P_{\text{bat}}^{\text{set}}(t)$ represent the required power values to be output by the engine and battery at time t . Because the energy supply needs to meet the constraints of supply and demand balance:

$$P_{\text{eng}}^{\text{set}}(t) + P_{\text{bat}}^{\text{set}}(t) = P_{\text{mot}}(t) \tag{25}$$

Here, $P_{\text{mot}}(t)$ is obtained from the energy flow graph, which is used to simulate the power of the traction motor. For the purpose of policy learning, the reward function based on energy efficiency comprehensively considers the dimensions of energy effectiveness, system stability and operating state responsiveness, and defines the total reward function as R_t, whose structure is expressed as:

$$R_t = -\alpha_1 \cdot \left| \text{SOC}(t) - \text{SOC}_{\text{ref}} \right| - \alpha_2 \cdot F_{\text{fuel}}(t) - \alpha_3 \cdot \Delta P_{\text{bat}}(t)^2 - \alpha_4 \chi_{\text{unstable}} \tag{26}$$

Among them, SOC_{ref} is the target range of SOC;

$F_{\text{fuel}}(t)$ represents the current engine fuel consumption; $\Delta P_{\text{bat}}(t)^2$ is the square term of the battery output power change, which is used to limit excessive instantaneous fluctuations; $\chi_{\text{unstable}}(t)$ is the penalty indicator for sudden working conditions, and $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are the weight coefficients of each item.

To verify the rationality of the trade-off relationship between the target items in the reward function, this paper conducts a systematic sensitivity analysis of the weight parameters. Under the premise of keeping other hyperparameters unchanged, three different weight combinations are set to evaluate their impact on the energy management system's overall performance, as shown in Table 3:

Table 3. Weight parameter evaluation.

Sequence	α_1	α_2	α_3	α_4	Average SOC fluctuation (%)	Average fuel consumption (L/100KM)	Standard deviation of power fluctuation (KW)	Emergency response time (S)
1	0.5	0.2	0.2	0.1	1.32	5.98	2.1	2.7
2	0.4	0.3	0.2	0.1	1.45	5.65	2.05	3
3	0.3	0.3	0.3	0.1	1.78	5.52	1.92	3.2

From Table 3, it can be seen that as α_1 increases, SOC stability improves, but fuel consumption slightly increases; an increase in α_3 reduces power fluctuations, but at the expense of SOC control effectiveness. Considering various indicators, the configurations with α_1 =0.4, α_2 =0.3, α_3 =0.2, and α_4 =0.1 demonstrate good balance and robustness in multiple testing scenarios.

Regarding algorithm implementation, the experience replay mechanism and target network structure are used to stabilize the learning process. In each stage of the interactive strategy, the four-tuple (S_t, A_t, R_t, S_{t+1}) is recorded in the replay pool, and the network weights are updated by small batch sampling; when calculating the required target, the target network $\hat{Q}(s,a;\theta^-)$ updated with a fixed step size is used to calculate the expected target [27-29]:

$$y_t = R_t + \gamma \max_{a'} \hat{Q}(S_{t+1}, a'; \theta^-) \quad (27)$$

Finally, the deep strategy is iteratively learned with the goal of minimizing the loss function.

$$\mathcal{L}(\theta) = \mathbb{E}_{(S_t, A_t, R_t, S_{t+1})}\left[\left(y_t - Q(S_t, A_t; \theta)^2\right)\right] \quad (28)$$

2) EMS Optimization

Due to the extremely complex state space after multi-sensor information fusion, DQN is prone to fall into local extreme values in the early stages of learning, resulting in delayed policy updates and inability to quickly adapt to sudden changes in data [30]. In a dynamic driving environment, vehicle conditions are changeable and energy flow patterns fluctuate dramatically. The basic DQN model is too conservative to adjust quickly, and the target Q value is updated with the maximum action value, which can easily lead to overestimation, thus affecting the vehicle's dynamic response and energy consumption optimization. To solve this problem, this paper introduces dual DQN and combines it with PER. By separating action selection and evaluation, the bias of overestimation is suppressed. By strengthening the weight of the critical point, the convergence and robustness of the algorithm in complex road environments are accelerated, as shown in Figure 3:

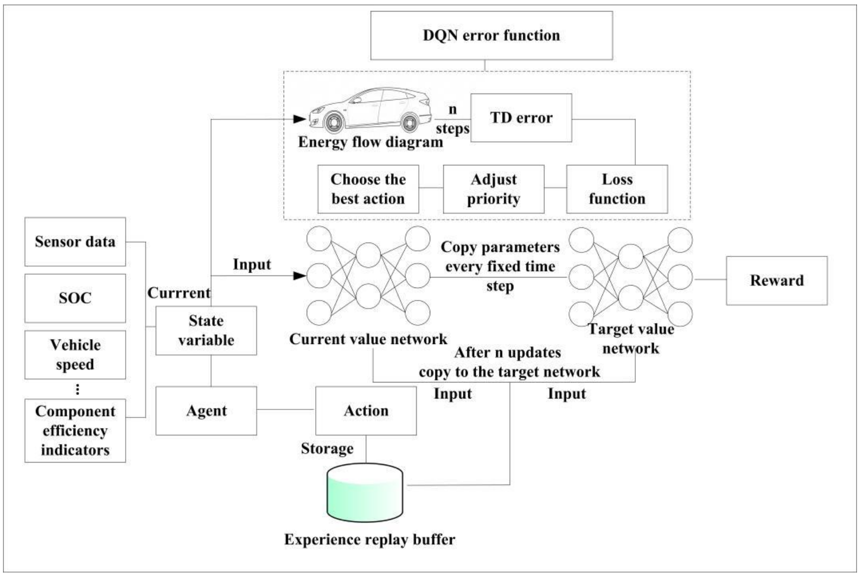

Figure 3. EMS optimization mechanism.

In dual DQN, the main network is used to select actions, and the target network evaluates the action value. The specific update is expressed as [31-33]:

$$y_i = r_i + \gamma Q_{\text{target}}\left(s_{i+1}, \arg\max_{a'} Q_{\text{online}}\left(S_{i+1}, a'; \theta\right); \theta^{-}\right) \quad (29)$$

Here, y_i is the target Q value at the i th iteration; r_i is the immediate reward; γ is the discount factor. Q_{online} and Q_{target} are the current value network and the target value network, and θ and θ^{-} are their corresponding parameters. By separating the action selection and evaluation processes, the over-estimation phenomenon is suppressed, and more reasonable energy flow decisions are generated under changing environments [34].

To enhance the automotive system's rapid response capability to complex working environments and sudden data, the PER mechanism is introduced to replace the conventional uniform sampling mode. In the traditional experience replay method, each sample is selected with equal probability, resulting in a lack of effective learning of critical mutation states. However, PER allocates sampling probability according to the importance of the sample, that is, through TD error. The larger the TD error, the greater the strategy error, and the more frequent updates are required. The specific sampling probability is defined as [35,36]:

$$P(i) = \frac{|\delta_i|\varsigma}{\sum_k |\delta_k|\varsigma} \quad (30)$$

Among them, δ_i is the TD error of the i th experience sample, and $\varsigma \in [0,1]$ is the factor that determines the priority. Through this mechanism, the deviation degree of key decision nodes can be quickly corrected to improve the adaptability to complex environmental conditions. To avoid introducing sampling bias, the gradient update is corrected together with the importance sampling weight ω_i:

$$\omega_i = \left(\frac{1}{N} \cdot \frac{1}{P(i)}\right)^{\beta} \quad (31)$$

In this process, as the training time goes by, the β value gradually increases to 1, thus ensuring the unbiasedness of the strategy in the later stage of algorithm training.

In the specific implementation, based on multi-source sensor information, the state vector of the system is dynamically constructed. Based on the system's dynamic constraints and energy balance conditions, a set of alternative action sets are generated, and the actions are executed according to the current strategy. At the end of each interaction, the experience is input into the priority experience replay pool, and the sampling probability is dynamically updated using the TD error. The target Q value is obtained according to the DoubleDQN principle, and the network parameters are iterated.

3. Energy Management Simulation Results and Analysis

A. Simulation Settings

To verify the effectiveness and robustness of the EMS under this method, a simulation analysis is performed. This paper combines the established sensor fusion model, dynamic model, and energy flow model to simulate the energy transfer between the components in the PHEV multi-sensor. The TensorFlow deep learning framework in Python can be used to establish the DQN training environment. The Real-World Drive Cycle Dataset in the FASTSim database is selected as the simulation data support source. The selected working condition data characteristics and their key indicators are shown in Table 4:

Table 4. Working condition data characteristics and their key indicators.

Working condition name	Working condition type	Maximum speed (km/h)	Average vehicle speed (km/h)	Total driving distance (km)
UDDS (Urban Dynamometer Driving Schedule)	Urban area	91.25	31.5	12.07
HWFET (Highway Fuel Economy Test)	High speed	97.7	77.7	16.45
US 06	High acceleration	129.2	80.3	12.86
NYCC (New York City Cycle)	Congested urban area	44.5	11.4	1.6

B. Algorithm Training

The dual DQN structure is based on a three-layer fully connected neural network, taking the state space S_t defined in this paper as the input dimension and the action space dimension, that is, the Q value of EMS, as the output. To enhance the algorithm's effectiveness and stability, the PER mechanism is combined to avoid the impact of Q value estimation oscillation on algorithm training. Considering the possible mutation interference under each working condition, $\varepsilon-$greedy strategy is adopted to dynamically adjust the exploration rate, which is initially set to 1.0 and eventually decays to 0.05. The key parameter configuration during DQN training is shown in Table 5:

Table 5. DQN parameter settings.

Model	Parameter	Specifications
Double DQN	Learning rate	0.0005
	Gamma	0.99
	Batch size	64
	Experience replay capacity	50,000
	Target network update cycle	Every 500 steps
	Initial/minimum value of ε	1.0 / 0.05
	ε decay rate	Linear decay to 0.05 every 5000 steps
	Maximum number of training epochs	1500 episodes

The basic DQN, Soft Actor-Critic (SAC), and Proximal Policy Optimization (PPO) are used for comparative experiments:

(1) Basic DQN: DQN without any improvement;

(2) SAC: based on the maximum entropy policy optimization principle, it has good performance in the continuous behavior space and is suitable for the smooth control requirements of the energy allocation strategy;

(3) PPO is a stabilization algorithm based on policy gradient and is widely used in optimal control problems.

Each algorithm is performed under the same training set and disturbance configuration. After the training, the convergence speed, energy consumption performance, response stability, computational efficiency dimension, and data mutation tolerance of the algorithm are quantitatively evaluated and analyzed.

C. Simulation Results

1) Convergence Speed

To verify the adaptability of different algorithms and the learning efficiency of strategies, 1500 rounds are run under the same training environment using the same initial parameters, perturbation ratio, and update mechanism. The convergence criterion is that the average Q value increase does not exceed 1% and lasts for more than 200 steps. The convergence steps and average fluctuation range of each algorithm under each working condition are shown in Figure 4:

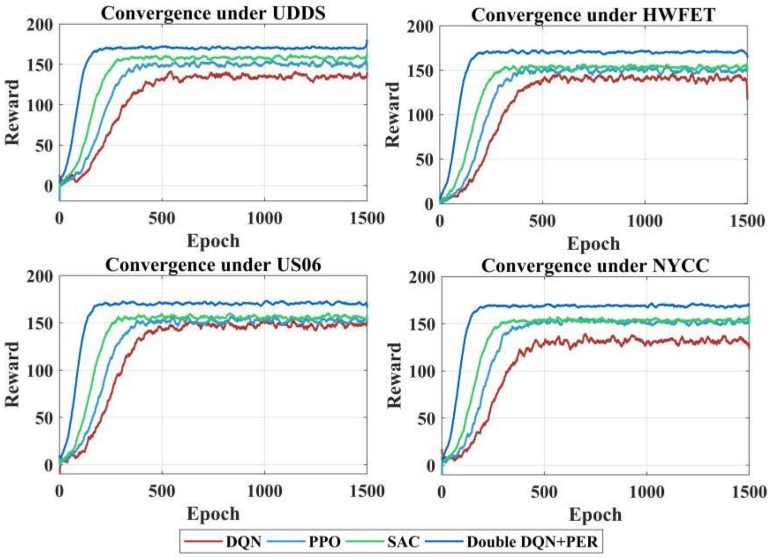

Figure 4. Convergence speed comparison.

From the convergence speed of Figure 4, it can be seen that the convergence advantage of the proposed algorithm under various industrial control systems is more significant. The reward curves of the algorithm in this paper all show rapid growth and reach a plateau around 400-600 rounds, and its convergence speed is higher than that of the basic DQN, PPO, and SAC algorithms. In contrast, the global convergence of the DQN algorithm is relatively slow, and the reward level after convergence has certain fluctuations, indicating that its generalization ability in complex states is not strong. As a policy gradient algorithm, PPO converges slightly faster than DQN, but overall, there are still problems of inconsistent convergence rate and easy to fall into local extreme values. In the initial stage, the learning efficiency of individuals in SAC is higher than that of PPO and DQN, and the rising slope of the reward curve is larger, showing the regulation effect on entropy change, but its convergence degree is still slightly lower than that of the algorithm in this paper.

The algorithm in this paper can overcome the overestimation problem that is common in traditional Q learning, and use dual networks to select and evaluate actions respectively, reduce policy deviations, and improve the stability and accuracy of learning. PER uses a weighted sampling method according to the size of the TD error, which increases the frequency of using samples that have a greater impact on learning and improves the algorithm's convergence efficiency, especially in the initial stage, where its training advantage is more significant.

2) *Energy Consumption Performance*

This paper uses 1000 time steps as the simulation cycle to simulate the change of battery and engine power over time, compares and analyzes the impact of different algorithms on system load size and output stability, and determines the energy saving effect and optimization performance of the algorithm under different operating conditions, as shown in Figures 5 and 6.

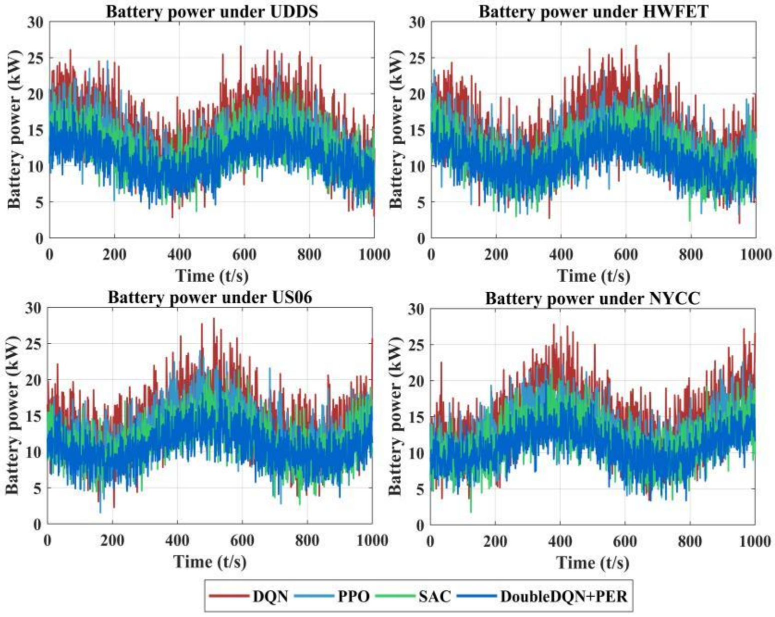

Figure 5. Battery energy consumption comparison.

From the results in Figure 5, it can be seen that the EMS algorithm in this paper has a significant advantage in energy consumption optimization. Under the four working conditions of Figure 5, its battery output power is relatively stable. The average power is controlled below 12 kW, which is lower than the other three algorithms and can effectively reduce the peak power. Especially in the case of frequent acceleration such as US06, the average power is only 11.22 kW. The algorithm in this paper adopts a double-layer network structure to solve the problem of overestimation and uses the PER mechanism to improve the learning efficiency of the critical point, the stability of battery energy management, and energy saving effect. In general, the proposed EMS considers both energy consumption and robustness of control strategies, and has strong adaptive capabilities.

In contrast, the battery power of the DQN algorithm fluctuates greatly under various working conditions, reflecting that its operating stability under complex load changes is not ideal. Although DQN has the ability of adaptive strategies, its learning process has slow convergence and is prone to local extreme value problems, which in turn causes large fluctuations in energy output and affects the system's energy efficiency and stability. The average battery power of the PPO algorithm EMS is above 13kW, which can effectively avoid sudden changes in system performance caused by sudden changes in strategy. However, it still has the problem of significant instantaneous power increase under complex operating conditions, indicating that its adaptability to rapid changes in operating conditions is still limited. The SAC power fluctuates between 11 and 13 kW, and at higher speeds, when the system frequently starts and stops for short periods of time, its average power level is higher than that of the EMS in this paper.

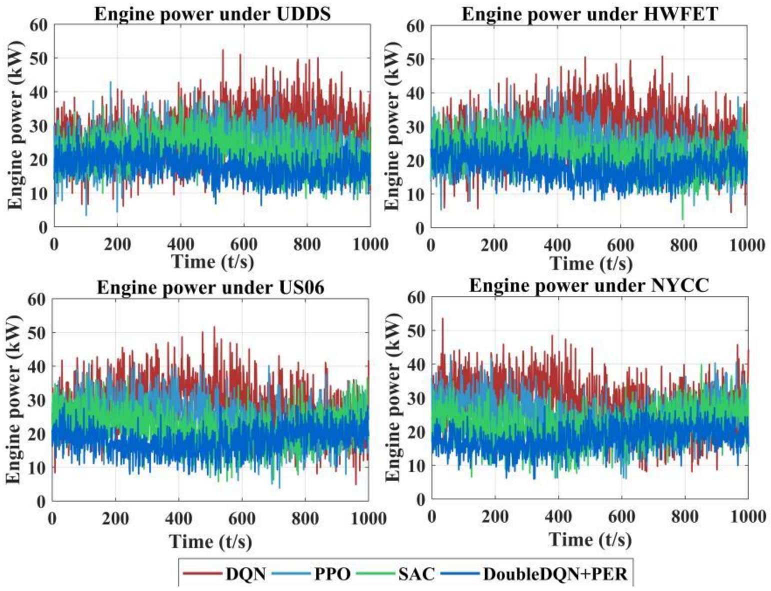

Figure 6. Engine energy consumption comparison.

Under the four operating conditions shown in Figure 6, the engine power outputs obtained by using different EMS algorithms are significantly different, reflecting the differences in energy allocation among various algorithms. According to the results of Figure 6, the EMS based on the algorithm in this paper shows the smallest and most stable engine power output in the whole time series. The average power output under the four working conditions does not exceed 19 kW, specifically 18.36 kW, 18.02 kW, 18.44 kW, and 17.93 kW. The variation is very small, and the curve is relatively smooth. This shows that the EMS algorithm in this paper tends to allocate more power to the battery, thereby reducing fuel consumption and engine losses. In comparison, the DQN algorithm has a higher power output, which is 28.53 kW, 28.12 kW, 29.41 kW, and 28.75 kW, respectively, with large fluctuations. This shows that its energy distribution mechanism is not stable, and it often needs to rely on the engine to cope with emergency driving needs, which causes the system's total energy consumption. The PPO and SAC algorithms are at an intermediate level, with moderate fluctuations, showing a certain degree of synergy, but their energy consumption levels are still higher than that of the algorithm in this paper. Compared with the other three methods, the EMS engine power curve under the algorithm in this paper is generally lower and can maintain good fluctuation stability under various operating conditions. The experimental results show that the algorithm in this paper can well implement the EMS system with battery priority and timely engine intervention. The dual DQN can well solve the problem of too high Q value of the traditional DQN algorithm, thereby improving the rationality of decision-making; PER can effectively utilize valuable experience and improve learning efficiency and strategy promotion. This combination enables the system to quickly grasp the critical point in operation during operation, thereby achieving a low-consumption effect.

3) *Response Stability*

To evaluate the response stability of each algorithm under different working conditions, the SOC control performance of the strategy under different dynamic load modes is analyzed. The overall fluctuation of the battery and the speed of change are characterized by two indicators: SOC fluctuation range and SOC change rate. The comparison results are shown in Figure 7:

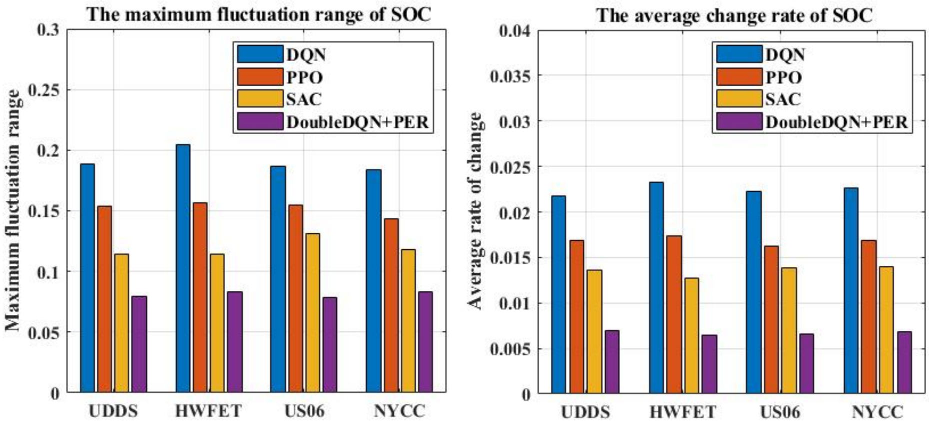

Figure 7. Response stability comparison.

According to the results shown in Figure 7, the proposed algorithm shows the best performance in terms of SOC control stability under UDDS, HWFET, US06, and NYCC conditions. Specifically, in terms of the SOC fluctuation range, the maximum fluctuation range of the proposed algorithm in the four working conditions does not exceed 0.09, which are 0.0792, 0.0829, 0.0784, and 0.0833, respectively, while the basic DQN is 0.1883, 0.2044, 0.1869, and 0.1836, respectively; the PPO algorithm is 0.1540, 0.1562, 0.1550, and 0.1435, respectively; the SAC algorithm is 0.1145, 0.1143, 0.1313, and 0.1182; the average SOC change rate of the proposed algorithm in the four working conditions does not exceed 0.007, which are 0.0069, 0.0064, 0.0066, and 0.0068, respectively; the basic DQN is 0.0217, 0.0233, 0.0222, and 0.0226, respectively; the PPO algorithm is 0.0169, 0.0174, 0.0163, and 0.0169, respectively; the SAC algorithm is 0.0136, 0.0127, 0.0138, and 0.0140. From the comparison results, the SOC fluctuation range and change rate of the proposed algorithm are significantly smaller than those of DQN, PPO and SAC algorithms.

The results show that the proposed method has better stability and better learning performance, which can effectively suppress SOC mutations and ensure the safe and efficient operation of the system. Its excellent performance comes from the structural optimization of the algorithm: the dual DQN adopts a two-layer network structure to reduce the error of excessive Q value estimation, making the strategy update more stable and accurate. On this basis, the PER mechanism is introduced to increase the weight of key samples, improve the sensitivity of the SOC upper and lower bound strategy convergence, and achieve stability control of the vehicle in a complex and changing environment.

To further evaluate the long-term stability of each algorithm, a simulation test scenario that runs continuously for 60 minutes is constructed in the simulation platform and divided into six 10-minute time periods (t1-t6). The difference between the maximum and minimum SOC values of the PHEV system in each time period is collected in each period, and the t1 period is used as the benchmark period. By expressing the increase ratio of the SOC fluctuation amplitude compared to the t1 period, the performance decay rate of EMS under each algorithm is calculated. The results are shown in Table 6:

Table 6. Performance decay rate.

Time interval	DQN(%)	SAC(%)	PPO(%)	Double DQN+PER(%)
t1	0	0	0	0
t2	8.5	3.7	3.8	1.3
t3	11	4.9	6.3	3.9
t4	13.4	6.2	7.5	5.2
t5	17.1	7.4	8.8	5.2
t6	19.5	8.6	11.3	6.5

From the results in Table 6, the long-term stability of the proposed algorithm is more ideal. The performance decay rate of the basic DQN algorithm is as high as 19.5% after 60 minutes, and the fluctuation increase is large, reflecting that its strategy is not stable enough over time, and there may be problems of overfitting training or poor strategy generalization ability. The performance decay rates of SAC and PPO after 60 minutes are 8.6% and 11.3%, respectively. The algorithm in this paper shows the smallest decay trend, and the performance decay rate after 60 minutes is only 6.5%, indicating that its experience priority sampling mechanism and target network structure effectively alleviate the strategy deviation and error accumulation, and have strong resistance to time disturbance.

4) Computational Efficiency

This paper compares the advantages and disadvantages of different algorithms in terms of EMS computational efficiency from two levels: single-step decision time and simulation run time. The single-step decision time is used to evaluate the average computational time required for the algorithm to output the control strategy. The total simulation run time is reflected in the overall computational overhead of the entire control loop. Under the same operating conditions and state input sequence, each algorithm runs independently, and their results are compared, as shown in Figure 8.

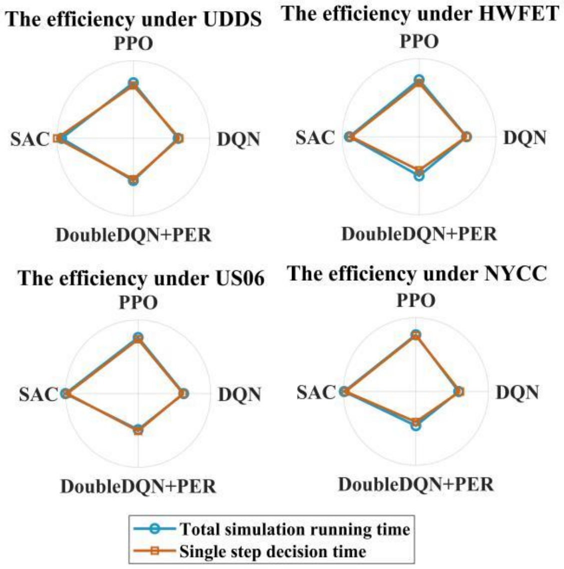

Figure 8. Comparison of computational efficiency.

From Figure 8, the single-step decision time and total simulation running time of the proposed algorithm under various working conditions are shorter than those of other algorithms. The single-step decision time of the algorithm in this paper is 0.0046 seconds, 0.0042 seconds, 0.0041 seconds, and 0.0039 seconds, respectively; the basic DQN is 0.0050 seconds, 0.0052 seconds, 0.0053 seconds, and 0.0050 seconds, respectively; the PPO algorithm is 0.0061 seconds, 0.0062 seconds, 0.0065 seconds, and 0.0065 seconds, respectively. The SAC algorithm is 0.0080 seconds, 0.0077 seconds, 0.0085 seconds, and 0.0084 seconds; in Figure 8A, Figure 8B, Figure 8C, and Figure 8D, the total simulation running time of the algorithm in this paper is 4.5559 seconds, 3.7107 seconds, 4.3609 seconds, and 3.5520 seconds, respectively; the basic DQN is 5.1926 seconds, 5.2906 seconds, 5.3302 seconds, and 5.2344 seconds, respectively; the PPO algorithm is 5.8693 seconds, 5.9391 seconds, 6.3524 seconds, and 6.5460 seconds, respectively; the SAC algorithm is 8.6139 seconds, 7.6572 seconds, 8.4201 seconds, and 8.3417 seconds.

From the results, after integrating PER, the algorithm in this paper can use historical data more effectively, speed up learning, and reduce computational costs. DoubleDQN reduces overestimation errors, improves learning stability, and accelerates convergence, thereby effectively shortening the strategy control cycle time. Through fast iterative training, an optimal control strategy function is obtained, which is the mapping relationship between the state and the corresponding optimal control action. In comparison, although the performance of the basic DQN and PPO algorithms is relatively stable, the algorithm complexity is high, and the solution speed is slow, especially when the state space dimension is high. SAC performs well under some working conditions, but due to its high computational cost, the algorithm’s overall performance is greatly limited. In summary, the method in this paper can effectively improve the computing speed of EMS and improve the energy management efficiency of the automotive system.

5) Data Mutation Fault Tolerance

To verify the data mutation fault tolerance of each algorithm under EMS, the abnormal conditions of the actual sensor are simulated, and two types of interference are introduced in each test condition:

(1) The random drift of the signal simulates the inaccurate output of low-precision sensors caused by high temperature and humidity;

(2) Data loss simulates the data loss of the Controller Area Network bus and randomly deletes 3-5 adjacent input signals.

The EMS fault tolerance performance of each algorithm after the introduction of abnormal sensor disturbance is evaluated. The engine load change rate (%/s) is used as a measurement indicator, and the average amplitude of the load value change per unit time is statistically analyzed to reflect the fault tolerance of EMS under each algorithm to sudden data anomalies. Among them,%/s represents the proportion of change in engine load per unit time (per second) in the case of signal random drift or data loss. The results are shown in Table 7 and Table 8:

Table 7. Load change rate under signal random drift interference.

Working condition	DQN(%/s)	SAC(%/s)	PPO(%/s)	Double DQN+PER(%/s)
UDDS	3.64	2.85	2.51	1.73
HWFET	4.21	3.06	2.82	2.01
US06	4.96	3.72	3.41	2.53
NYCC	4.68	3.47	3.29	2.34
Mean	4.37	3.28	3.01	2.15

Table 8. Load change rate under data loss.

Working condition	DQN(%/s)	SAC(%/s)	PPO(%/s)	Double DQN+PER(%/s)
UDDS	3.79	3.08	2.76	1.95
HWFET	4.29	3.22	3.00	2.14
US06	5.07	4.03	3.64	2.69
NYCC	4.93	3.79	3.54	2.52
Mean	4.52	3.53	3.24	2.33

From the results in Table 7 and Table 8, it can be seen that both the random drift of sensor signals and the loss of data have a certain impact on the PHEV engine load. Under the two interference conditions, the average load change rate of the basic DQN is 4.37%/s and 4.52%/s, showing that it has a weak ability to adapt to emergencies and is prone to significant fluctuations during the control process. Although SAC (3.28%/s, 3.53%/s) and PPO (3.01%/s, 3.24%/s) have been improved compared to the basic DQN, they still have a large jitter problem under high dynamic conditions such as US06, and there are still hidden dangers in stability. The method in this paper has good fault tolerance for both types of interference. The average change rate under random signal drift is 2.15%/s, and the average change rate under data loss is 2.33%/s, which is significantly better than other methods and shows strong fault tolerance. This paper effectively suppresses decision instability caused by overestimation of quantization through the dual DQN structure, and uses the PER mechanism to achieve learning and memory of mutation points, thereby improving the generalization ability of the strategy and providing a more reliable guarantee for multi-sensor energy management of PHEV under actual road conditions.

Although strategy optimization algorithms such as SAC and PPO can also achieve effective energy management, they typically require more complex model parameter adjustments and may face computational efficiency issues in high-dimensional state spaces. Double DQN+PER utilizes the advantages of deep learning by dynamically adjusting sampling probabilities to improve sample utilization and convergence speed, especially suitable for real-time online updates. In dealing with sudden changes in sensor data and complex road conditions, the "dual DQL+PER" architecture exhibits stronger robustness and adaptability. Due to the PER mechanism's ability to dynamically adjust the importance of empirical samples based on TD errors, the model focuses more on state transitions that are crucial for policy improvement, thereby enhancing the system's anti-interference ability.

6) Comparison of Advanced Methods

To further highlight the effectiveness of this paper, the performance of the "dual DQN+PER" architecture and two cutting-edge methods, namely MPC based on RL hybrid framework and energy management based on neural differential equations, are compared in extreme scenarios of severe congestion in urban traffic and complex road conditions in mountainous areas. The results are shown in Table 9:

Table 9. Comparison of extreme scenarios.

Scene	Model	SOC fluctuation range	Average power output (kW)	Energy consumption change rate (%)	Response time (s)
Severe congestion	Double DQN+PER	0.082	22.5	6.2	0.563
	RL Hybrid MPC	0.136	23.3	7.3	0.711
	Neural differential equation	0.114	23.5	7.5	0.832
Complex road conditions in mountainous areas	Double DQN+PER	0.073	19.2	5.8	0.414
	RL Hybrid MPC	0.095	21.5	6.5	0.608
	Neural differential equation	0.118	22.7	7.0	0.712

From Table 9, it can be seen that "dual DQN+PER" exhibits the smallest SOC fluctuation range, lower average power output, better energy consumption change rate, and faster system response time in the test scenario. This indicates that the method has significant advantages in dealing with sensor noise and responding quickly to external environmental changes. Especially in dealing with extreme scenarios, the stability and adaptability of "dual DQN+PER" are superior to the other two methods, which can better maintain system performance.

4. Conclusions

The sudden changes in existing sensor data and complex operating conditions result in poor adaptive and stable performance of PHEV multi-sensor EMS. To improve the ability of PHEVs to cope with unstable interference and sensor anomalies, this paper proposes a dual DQN+PER architecture based on reinforcement learning algorithm for optimizing the multi-sensor energy management system of PHEVs. This system not only improves the energy efficiency of the vehicle itself, but also promotes seamless integration with renewable energy networks. Through precise energy scheduling and dynamic response mechanisms, PHEVs can quickly adapt to external environmental changes under non-steady state conditions. This flexibility makes PHEVs an important distributed energy storage unit in smart grids, helping to balance supply and demand and improve the entire power system's stability. Under complex working conditions, the performance of this paper's EMS under the complex operating conditions of UDDS, HWFET, US06, and NYCC is analyzed. Experiments have shown that compared with the basic DQN, SAC, and PPO algorithms, the EMA under the algorithm in this paper is more ideal in convergence speed, energy consumption performance, response stability, computational efficiency dimension, and data mutation tolerance. The average power output of the battery and engine under four complex working conditions is controlled within 12kW and 19kW, respectively. The maximum fluctuation range of State of Charge (SOC) does not exceed 0.09; the average change rate does not exceed 0.007; the performance decay rate after 60 minutes is only 6.5%; the average change rate under signal random drift and data loss is 2.15%/s and 2.33%/s, respectively. It can quickly converge to stability within 400-600 rounds and has excellent performance in battery, engine power control, and SOC fluctuation suppression. It can effectively reduce peak power and power consumption, improve overall energy efficiency and stability, help PHEVs achieve anti-interference under complex working conditions, and maintain good decision consistency when sensors have abnormalities such as analog signal drift and frame loss. This study provides a certain basis for the energy management of PHEV under complex working conditions and interference conditions, but there are still some shortcomings. In the construction of EMS, the multi-objective optimization problem has not been deeply explored. Although this paper does not consider external environmental factors such as road slope and ambient temperature, it has not yet delved into the specific impacts under different weather conditions. Weather conditions such as temperature, humidity, and precipitation not only affect the efficiency of the vehicle's power system but may also indirectly affect battery performance and energy consumption. Future research should further analyze the specific impacts of various weather conditions on PHEV energy management, integrate online learning mechanisms with multi-agent collaborative decision-making, and expand its adaptability and intelligence on a larger scale.

Acknowledgment

None

Consent to Publish

The manuscript has neither been previously published nor is under consideration by any other journal. The authors have all approved the content of the paper.

Funding

None

References

[1] X.J. Zeng, X.Q. Sun, F. Zhao. Energy-saving intelligent manufacturing optimization scheme for new energy vehicles. International Journal of Emerging Electric Power Systems, 2022, 23(6), 913-926. DOI: 10.1515/ijeeps-2022-0127

[2] Isabel C. Gil-Garcia, M. Socorro Garcia-Cascales, H. Dagher, A. Molina-Garcia. Electric vehicle and renewable energy sources: Motor fusion in the energy transition from a multi-indicator perspective. Sustainability, 2021, 13(6), 3430-3448. DOI: 10.3390/su13063430

[3] C. Yang, M.J. Zha, W.D. Wang, K.J. Liu, C.L. Xiang. Efficient energy management strategy for hybrid electric vehicles/plug-in hybrid electric vehicles: review and recent advances under intelligent transportation system. IET Intelligent Transport Systems, 2020, 14(7), 702-711. DOI: 10.1049/iet-its.2019.0606

[4] F.Q. Zhang, L.H. Wang, S. Coskun, H. Pang, Y.H. Cui, J.Q. Xi. Energy management strategies for hybrid electric vehicles: Review, classification, comparison, and outlook. Energies, 2020, 13(13), 3352-3387. DOI: 10.3390/en13133352

[5] A.Y. Cheng, Y. Xin, H. Wu, L.X. Yang, B.H. Deng. A review of sensor applications in electric vehicle thermal management systems. Energies, 2023, 16(13), 5139-5167. DOI: 10.3390/en13133352

[6] K. Purohit, S. Srivastava, V. Nookala, V. Joshi, P. Shah, R. Sekhar, et al. Soft sensors for state of charge, state of energy, and power loss in formula student electric vehicle. Applied System Innovation, 2021, 4(4), 78-104. DOI: 10.3390/asi4040078

[7] A. Mousaei, Y. Naderi, I.S. Bayram. Advancing state of charge management in electric vehicles with machine learning: A technological review. IEEE Access, 2024, 12, 43255-43283. DOI: 10.1109/ACCESS.2024.3378527

[8] X. Wang, S. Wang, X.X. Liang, D.W. Zhao, J.C. Huang, X. Xu. Deep reinforcement learning: A survey. IEEE Transactions on Neural Networks and Learning Systems, 2022, 35(4), 5064-5078. DOI: 10.1109/TNNLS.2022.3207346

[9] R.X. Zhang, C. Huang, M.M. Wang. Research Status and Development Trend of Energy Management Strategy for Hybrid Electric Vehicles. Forestry Machinery and Woodworking Equipment, 2022, 50(10), 50-55. DOI: 10.13279/j.cnki.fmwe.2022.0151

[10] Y.Z. Zhu, X.Y. Li, Q. Liu, S.H. Li, Y. Xu. A comprehensive review of energy management strategies for hybrid electric vehicles. Mechanical Sciences, 2022, 13(1), 147-188. DOI: 10.5194/ms-13-147-2022, 2022

[11] I. Kranthikumar, C.H. Srinivas, T. Vamsee Kiran, P. Pradeep, V. Balamurugan. A novel hybrid approach for efficient energy management in battery and supercapacitor based hybrid energy storage systems for electric vehicles. Electrical Engineering, 2025, 107(1), 1-17. DOI: 10.1007/s00202-024-02483-9

[12] C.W. Hong, Q.H. Liu, Y.B. Zhang. Real-time energy management optimization strategy for electric vehicles based on LSTM network learning. Power Demand Side Management, 2021, 23(3), 13-18. DOI: 10.3969/j.issn.1009-1831.2021.03.004

[13] X.Y. An, Y.F. Li, J.B. Sun, S.N. Sun, Y.P. Shen. Energy management strategy of electric vehicle dual-source hybrid energy storage system based on fuzzy logic. Power System Protection and Control, 2021, 49(16), 135-142. DOI: 10.19783/j.cnki.pspc.201266

[14] J.J. Guo, Y. Wang, D.P. Shi, F.L. Chu, J.H. Wang, Z.L. Lv. Comparative Study and Optimization of Energy Management Strategies for Hydrogen Fuel Cell Vehicles. World Electric Vehicle Journal, 2024, 15(9), 414-433. DOI: 10.3390/wevj15090414

[15] T. Liu, W.W. Huo, B. Lu, J.W. Li. Reinforcement Learning-Based Co-Optimization of Adaptive Cruise Speed Control and Energy Management for Fuel Cell Vehicles. Energy Technology: Generation, Conversion, Storage, Distribution, 2024, 12(1), 2-12. DOI: 10.1002/ente.202300541

[16] A. Zare, M. Boroushaki. A knowledge-assisted deep reinforcement learning approach for energy management in hybrid electric vehicles. Energy, 2024, 313(30), 1-10. DOI: 10.1016/j.energy.2024.134113

[17] X.Y. Li, H.W. He, J.D. Wu. Knowledge-Guided Deep Reinforcement Learning for Multiobjective Energy Management of Fuel Cell Electric Vehicles. IEEE Transactions on Transportation Electrification, 2025, 11(1), 2344-2355. DOI: 10.1109/TTE.2024.3421342

[18] B.W. Zou, B.F. Zhang, Q.H. Ling, Y. Lian, J. Liu, S.Z. Du, et al. Research on energy management strategy of extended-range new energy vehicles based on reinforcement learning. Journal of Southwest University (Natural Science Edition), 2022, 44(3), 2-11. DOI: 10.13718/j.cnki.xdzk.2022.03.001

[19] Z.Y. Chen, Z.Y. Fang, R.X. Yang, Q.Q. Yu, M.Q. Kang. Energy management strategy of hybrid electric vehicle based on deep reinforcement learning. Transactions of China Electrotechnical Society, 2022, 37(23), 6157-6168. DOI: 10.19595/j.cnki.1000-6753.tces.211342

[20] X.L. Tang, J.X. Chen, T. Liu, J.C. Li, X.S. Hu. Research on intelligent following control and energy management strategy of hybrid electric vehicle based on deep reinforcement learning. Journal of Mechanical Engineering, 2021, 57(22), 237-246. DOI: 10.3901/JME.2021.22.237

[21] A. Ahmadian, B. Mohammadi-Ivatloo, A. Elkamel. A review on plug-in electric vehicles: Introduction, current status, and load modeling techniques. Journal of Modern Power Systems and Clean Energy, 2020, 8(3), 412-425. DOI: 10.35833/MPCE.2018.000802

[22] J. Oncken, B. Chen. Real-time model predictive powertrain control for a connected plug-in hybrid electric vehicle. IEEE Transactions on Vehicular Technology, 2020, 69(8), 8420-8432. DOI: 10.1109/TVT.2020.3000471

[23] X.L. Tang, T. Jia, X.S. Hu, Y.J. Huang, Z.W. Deng, H.Y.

Pu. Naturalistic data-driven predictive energy management for plug-in hybrid electric vehicles. IEEE Transactions on Transportation Electrification, 2020, 7(2), 497-508. DOI: 10.1109/TTE.2020.3025352

[24] A. Mousaei, Y. Naderi. Predicting Optimal Placement of Electric Vehicle Charge Stations Using Machine Learning: A Case Study in Glasgow, UK. 2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG). IEEE, 2025. DOI: 10.1109/ICREDG66184.2025.10966078

[25] Q. Zhou, D.Z. Zhao, B. Shuai, Y.F. Li, H. Williams, H.M. Xu. Knowledge implementation and transfer with an adaptive learning network for real-time power management of the plug-in hybrid vehicle. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(12), 5298-5308. DOI: 10.1109/TNNLS.2021.3093429

[26] E. Ehsani, K.V. Singh, H.O. Bansal, R. Tafazzoli Mehrjardi. State of the art and trends in electric and hybrid electric vehicles. Proceedings of the IEEE, 2021, 109(6), 967-984. DOI: 10.1109/JPROC.2021.3072788

[27] D.Y. Guo, G.Y. Lei, H.C. Zhao, F. Yang, Q. Zhang. Quadruple Deep Q-Network-Based Energy Management Strategy for Plug-in Hybrid Electric Vehicles. Energies, 2024, 17(24), 6298-6317. DOI: 10.3390/en17246298

[28] H.C. Liu, H.L. Wang, M. Yu, Y.L. Wang, Y. Luo. Long Short-Term Memory–Model Predictive Control Speed Prediction-Based Double Deep Q-Network Energy Management for Hybrid Electric Vehicle to Enhanced Fuel Economy. Sensors, 2025, 25(9), 2784-2816. DOI: 10.3390/s25092784

[29] A. Mousaei, Y. Naderi, S. Mekhilef, S. GOLESTAN, A. Iqbal. Optimal Placement of Electric Vehicle Charging Stations Using Machine Learning: A Comprehensive Review. Available at SSRN, 2025. DOI: 10.2139/ssrn.5195167

[30] C. Montaleza, P. Arevalo, J. Gallegos, F. Jurado. Enhancing energy management strategies for extended-range electric vehicles through deep Q-learning and continuous state representation. Energies, 2024, 17(2), 514-534. DOI: 10.3390/en17020514

[31] X.Y. Fan, L.L. Guo, J.L. Hong, Z.H. Wang, H. Chen. Constrained hierarchical hybrid Q-network for energy management of HEVs. IEEE Transactions on Transportation Electrification, 2024, 10(4), 9579-9590. DOI: 10.1109/TTE.2024.3353765

[32] L.J. Han, X. Zhou, N.K. Yang, H. Liu, L. Bo. Multi-objective energy management for off-road hybrid electric vehicles via nash DQN. Automotive Innovation, 2025, 8(1), 140-156. DOI: 10.1007/s42154-024-00323-x

[33] A. Mousaei. Analyzing locational inequalities in the placement of electric vehicle charging stations using machine learning: A case study in Glasgow. Next Research, 2025, 2(1), 100123. DOI: 10.1016/j.nexres.2024.100123

[34] Z.T. Pang, J.Q. Zhang, X.H. Jiao, L. Ren. Energy management strategy combining double deep Q-networks and demand torque prediction for connected hybrid electric vehicles. IEEJ Transactions on Electrical and Electronic Engineering, 2024, 19(2), 234-246. DOI: 10.1002/tee.23942

[35] H. Li, X.Z. Qian, W. Song. Prioritized experience replay based on dynamics priority. Scientific Reports, 2024, 14(1), 6014-6021. DOI: 10.1038/s41598-024-56673-3

[36] Z.W. Lou, Y.Y. Wang, S. Shan, K.J. Zhang, H.K. Wei. Balanced prioritized experience replay in off-policy reinforcement learning. Neural Computing and Applications, 2024, 36(25), 15721-15737. DOI: 10.1007/s00521-024-09913-6